Artificial Intelligence

Artificial Neural Network

Pham Viet Cuong
Dept. Control Engineering & Automation, FEEE
Ho Chi Minh City University of Technology

Artificial Neural Networks

- ✓ Computing systems inspired by biological neural networks
- ✓ Consists of several processing elements that receive inputs and deliver outputs
- ✓ "Learn" to perform tasks by considering examples
- ✓ Be able to model nonlinear processes

Artificial Neural Networks

- Applications:
  - Natural Language Processing
    - Text classification
    - Language generation
    - Document summarization
    - Machine translation
    - Speech recognition
    - Character recognition
    - Spell checking

Artificial Neural Networks

- Applications:
  - System identification and control
    - Vehicle control
    - Process control
    - Auto-piloting
    - Autonomous driving cars
    - Robot navigation
    - Component fault detection and simulation
    - Chip failure analysis
    - Trajectory prediction
    - Petroleum exploration

Artificial Neural Networks

- Applications:
  - Time series forecasting and prediction
    - Weather forecasting
    - Stock market prediction
    - Cost models
  - Medical
    - Cancers detection
    - Medical image analysis
    - EEGs
    - Drug design

Artificial Neural Networks

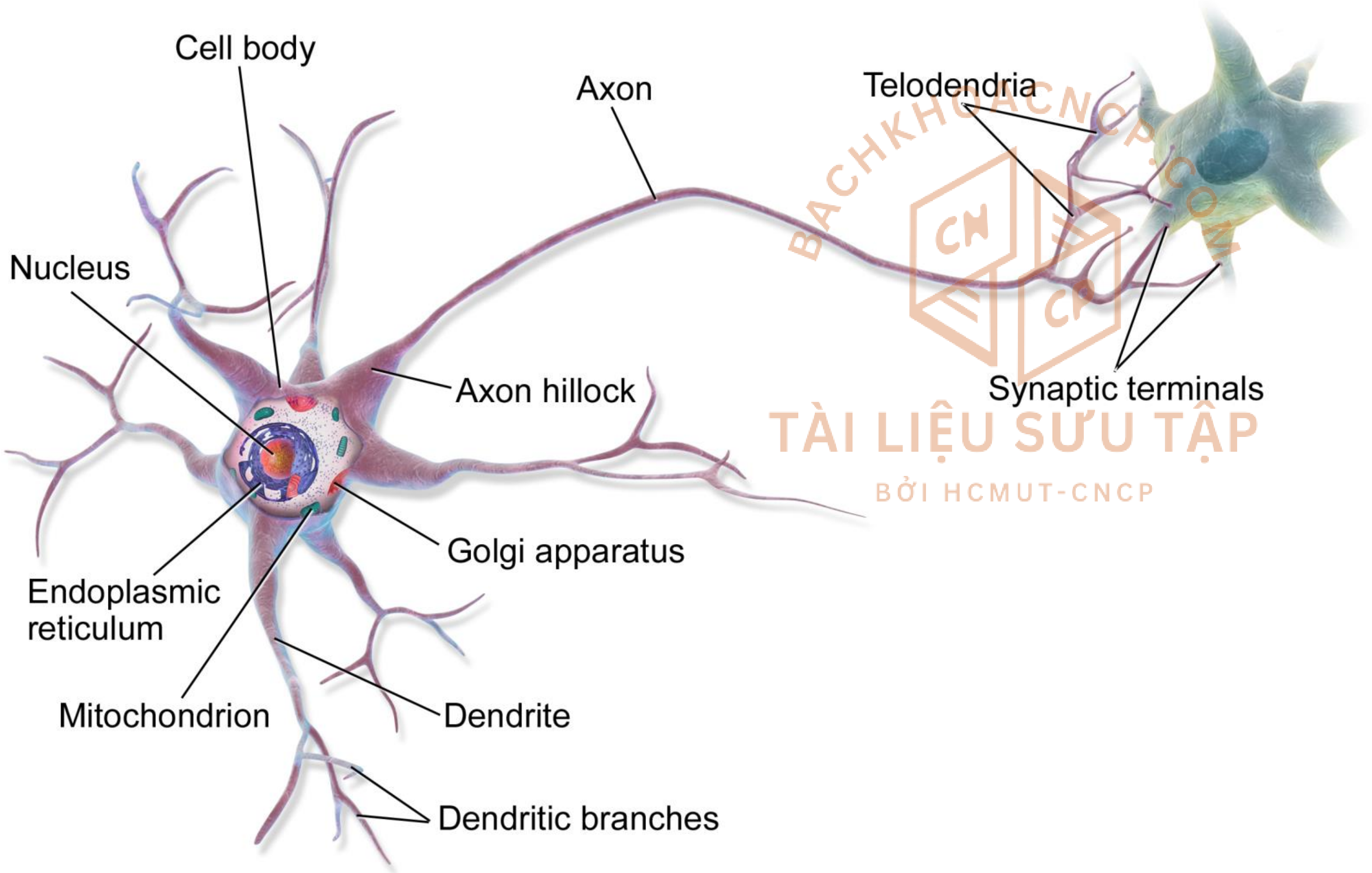
- Applications:
  - Image/video/audio analysis
    - Image recognition
    - Image compression
    - Radar and sonar image classification
    - Signature verification

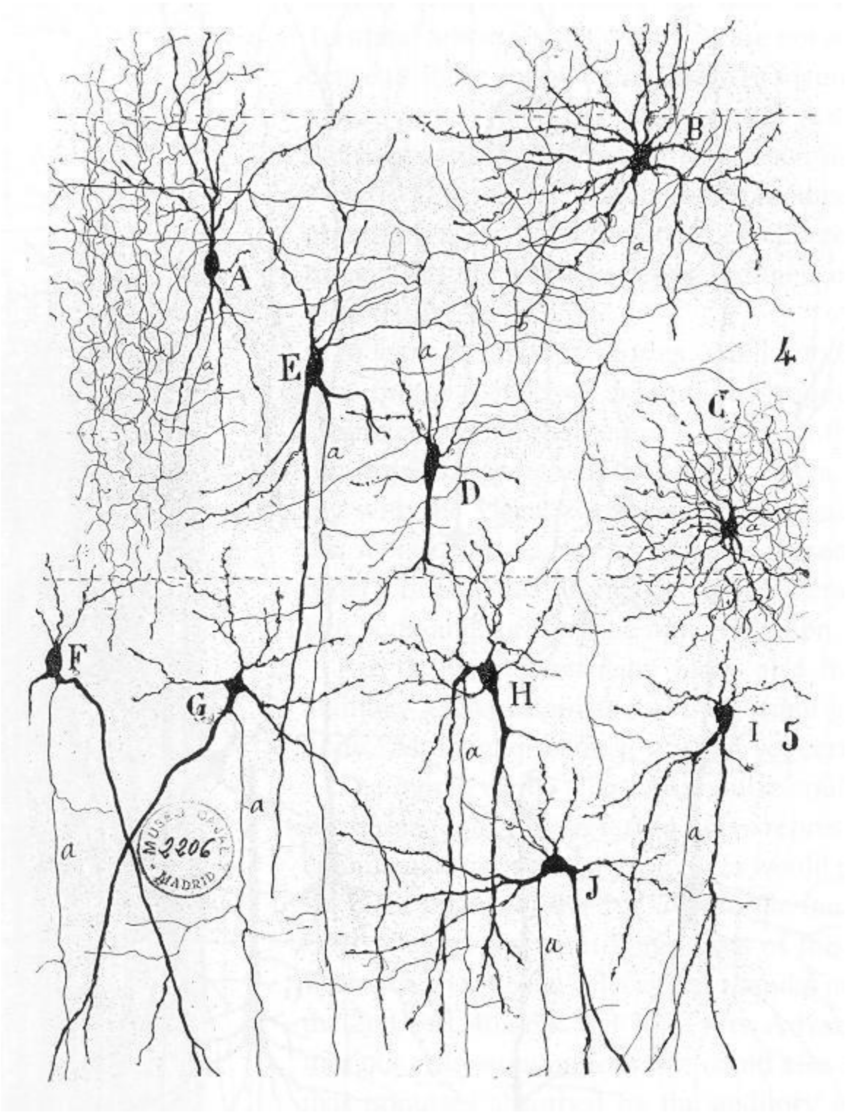
Artificial Neural Networks

- Applications:
  - Education:
    - Adaptive learning software
    - Student performance modeling
    - Personality profiling
  - Bussiness analytics
    - Customer behavior modeling
    - Customer segmentation
    - Market research
  - Banking
    - Credit and loan application evaluation
    - Fraud and risk evaluation

Artificial Neural Networks

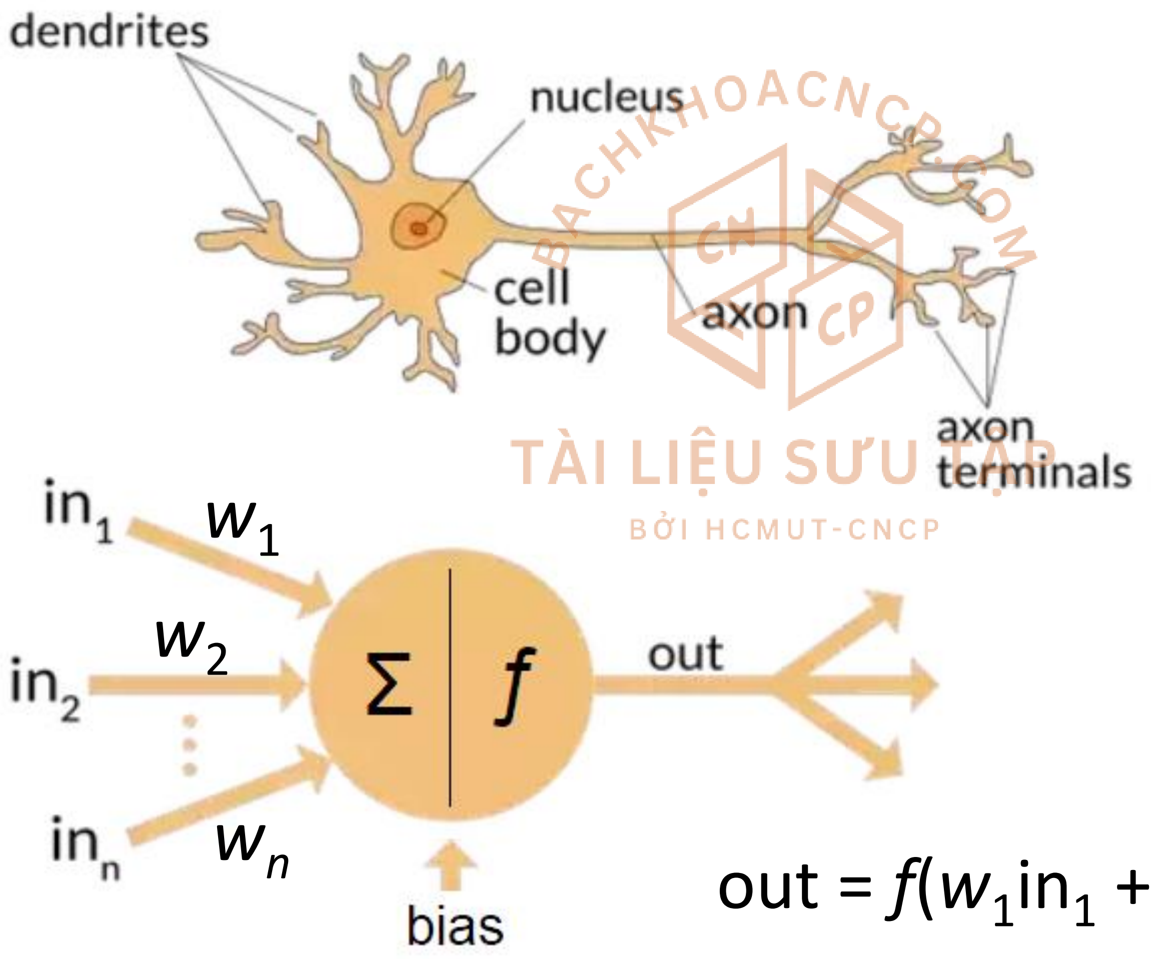
✓ Biological neural network





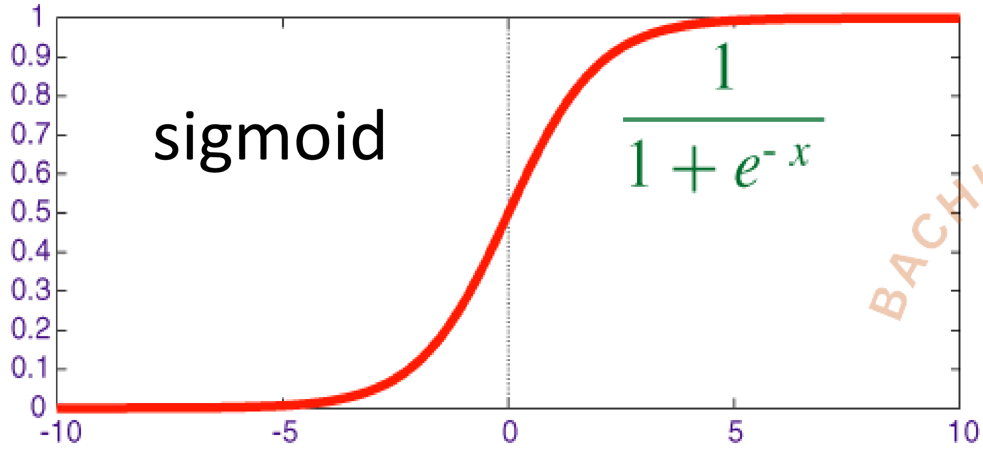
Artificial Neural Networks

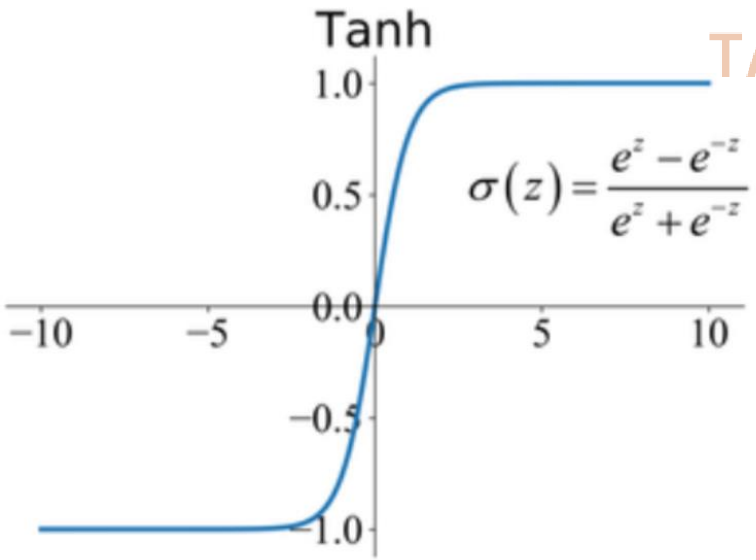
- ✓ Artificial neural network

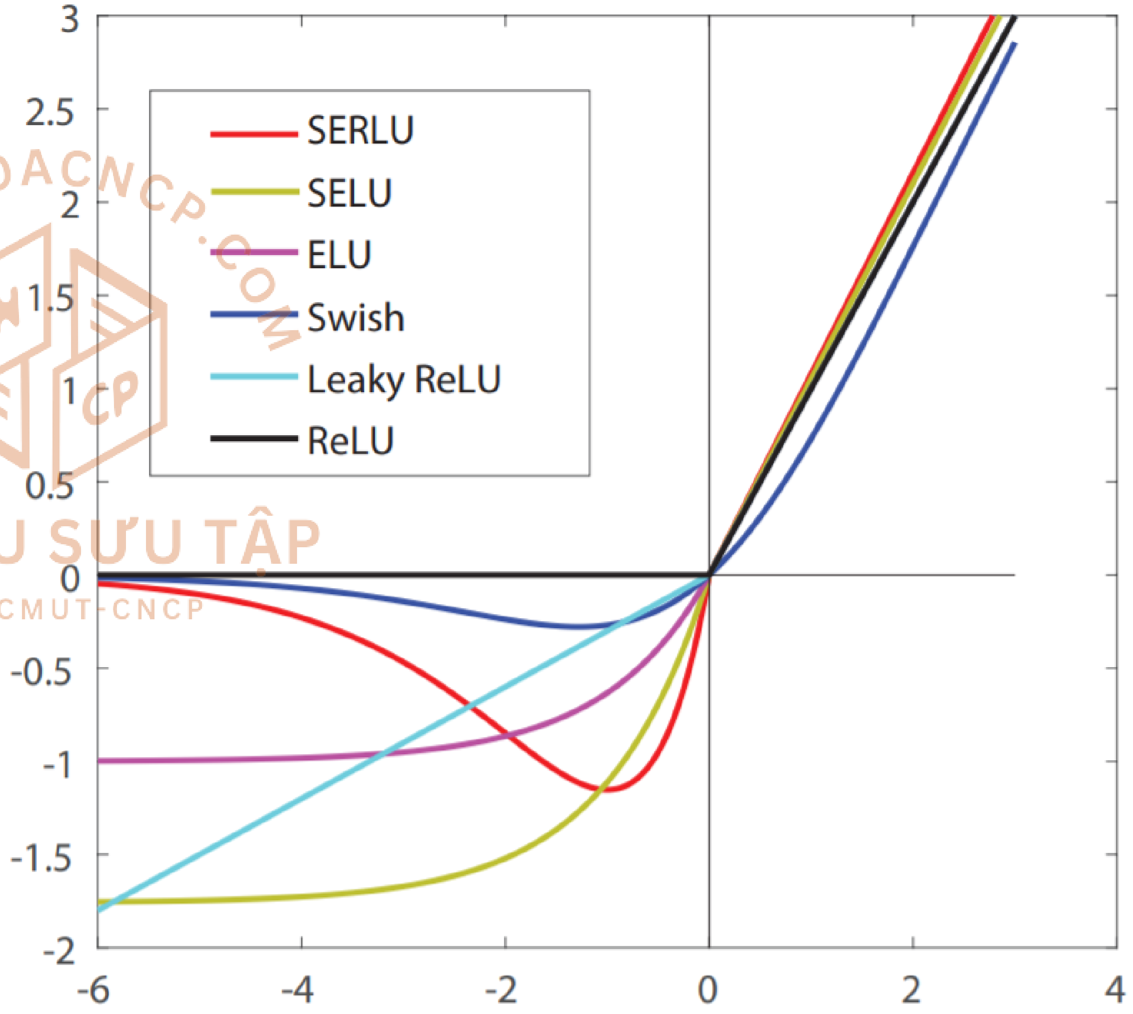


$$out = f(w_1 in_1 + w_2 in_2 + \ldots + w_n in_n + bias)$$

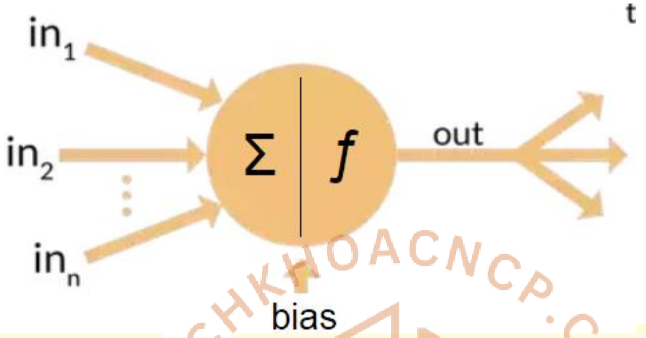
✓ Activation function

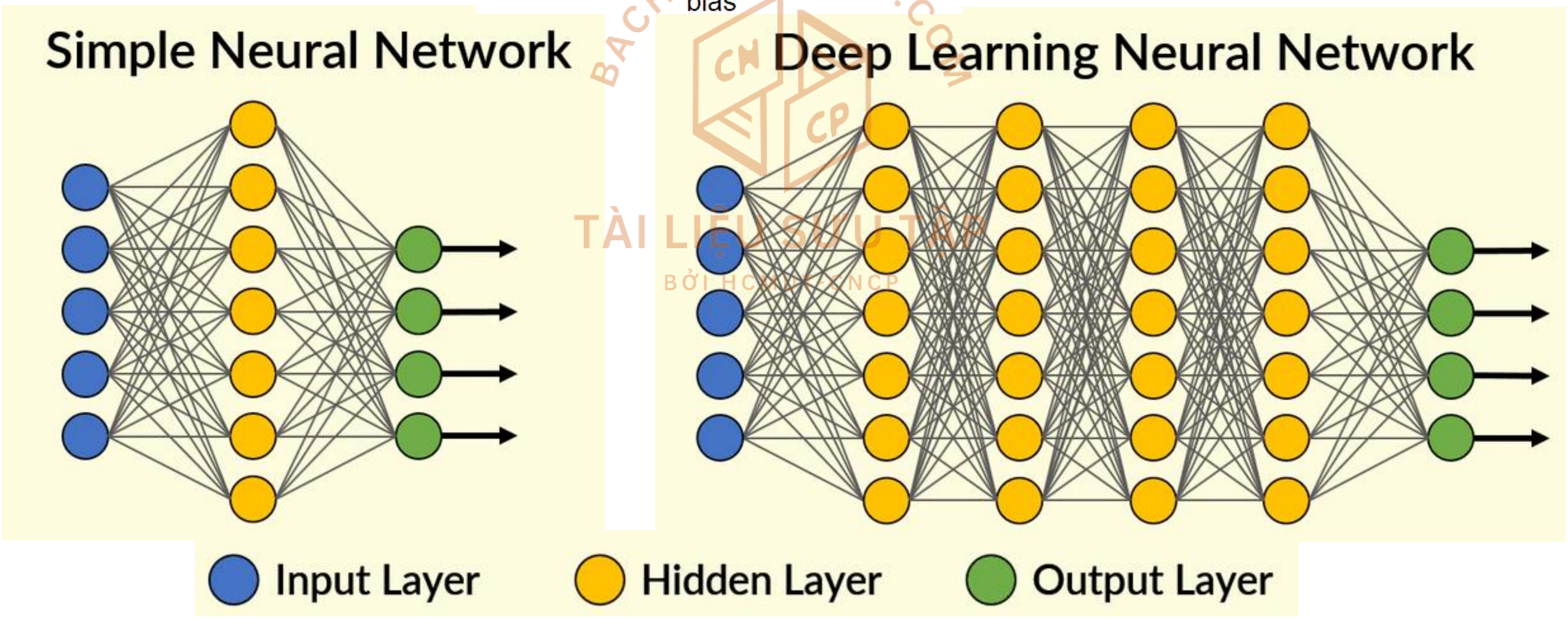






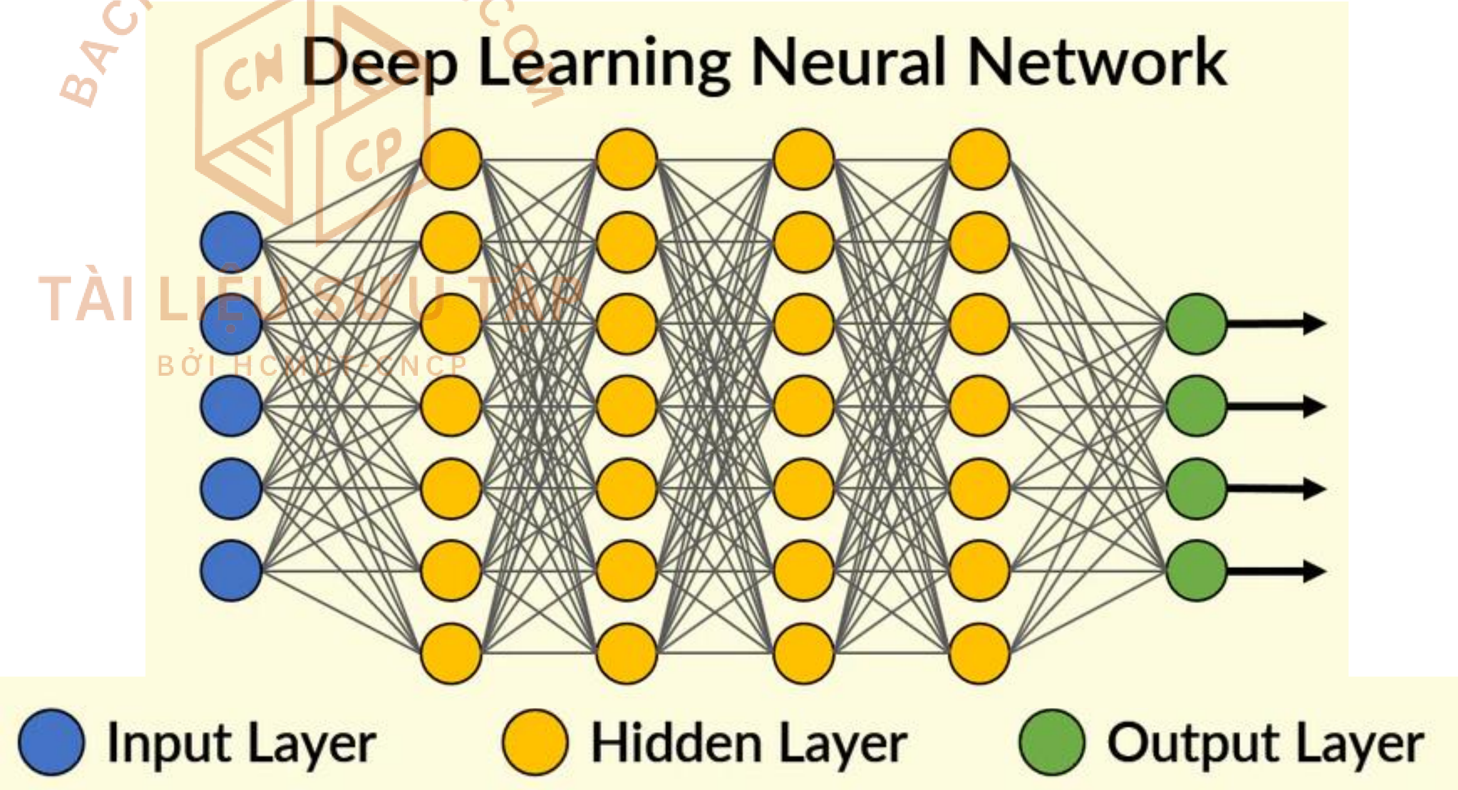
Artificial Neural Networks





Artificial Neural Networks

- Example: self-driving car
  - Input (raw data/features)
  - Output (steering angle, throttle, brake)
  - Hidden layers



Artificial Neural Networks

- ✓ Why nonlinear activation function?
- ✓ Why bias?

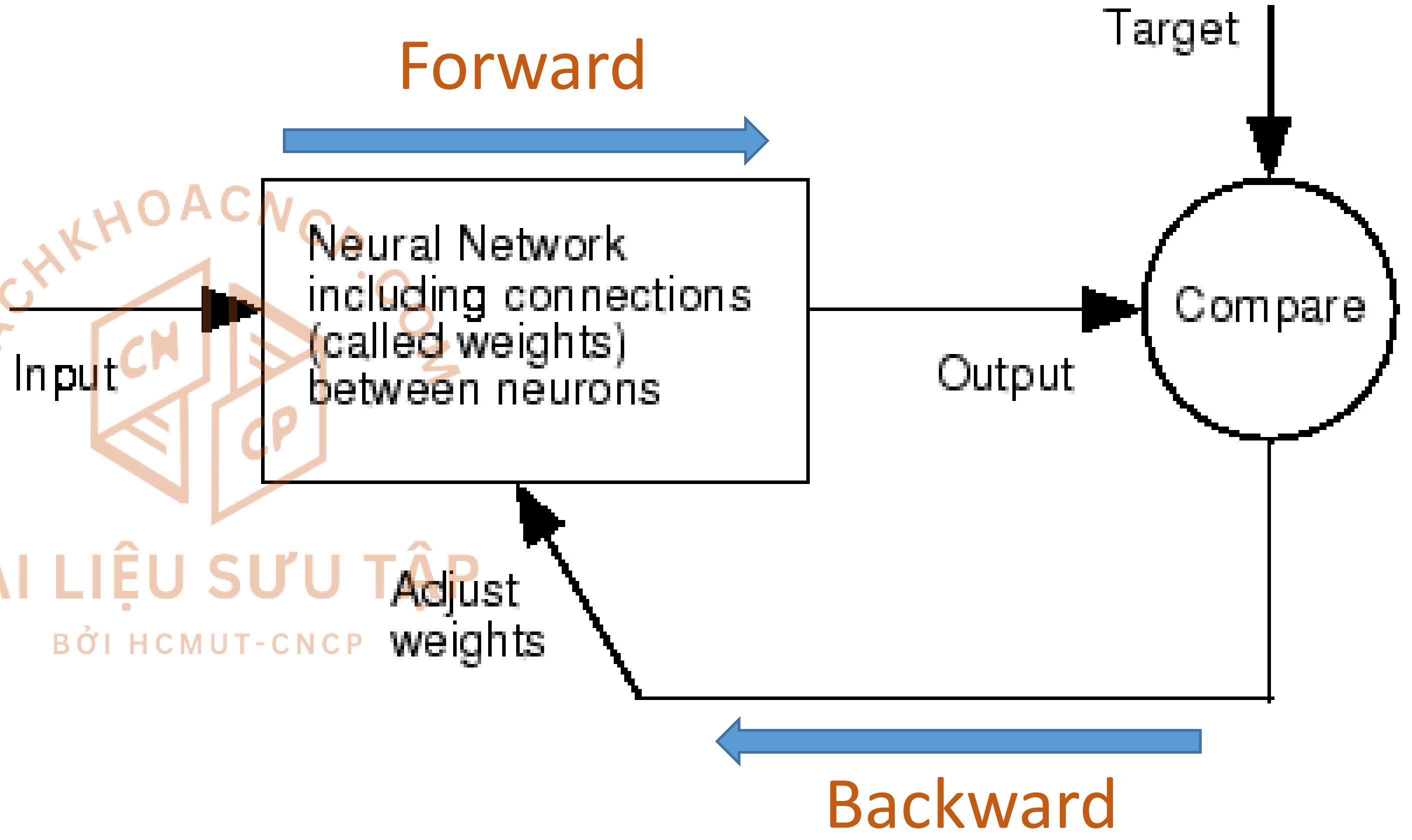
Artificial Neural Networks

✓ Useful links

- Link 1
- Link 2
- Link 3

Artificial Neural Networks

- ✓ Error back propagation
  - Framework



  - Loss function:
    - Squared L^2 norm
    - Angle between 2 vectors
    - Cross entropy

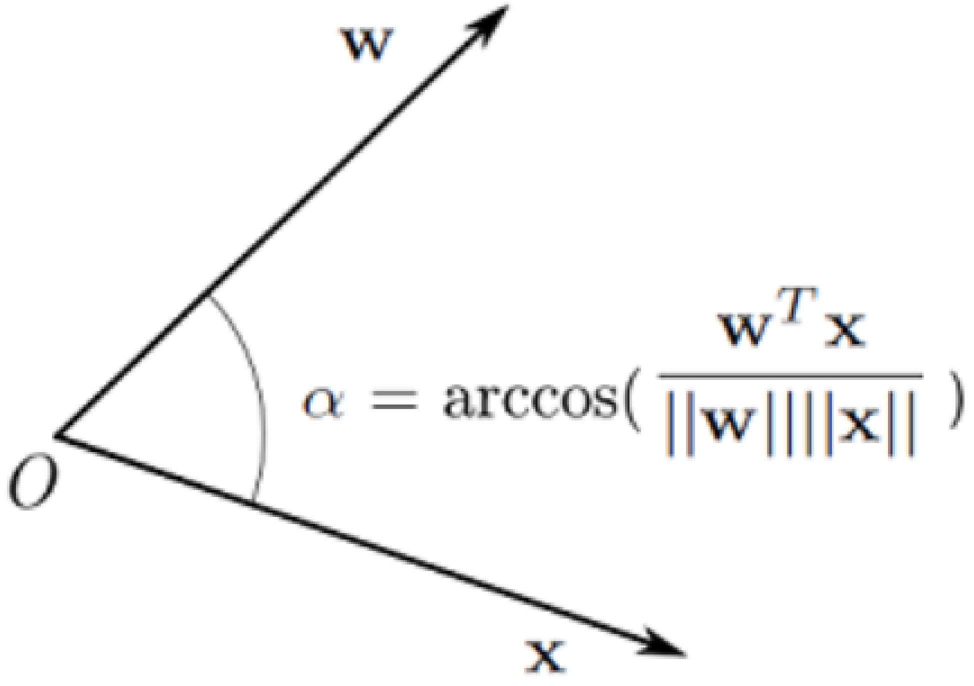
✓ Error back propagation

❖ Loss function:

➢ Euclidean norm (2-norm)

$$\|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$$

➢ Angle between 2 vectors



➢ Cross entropy

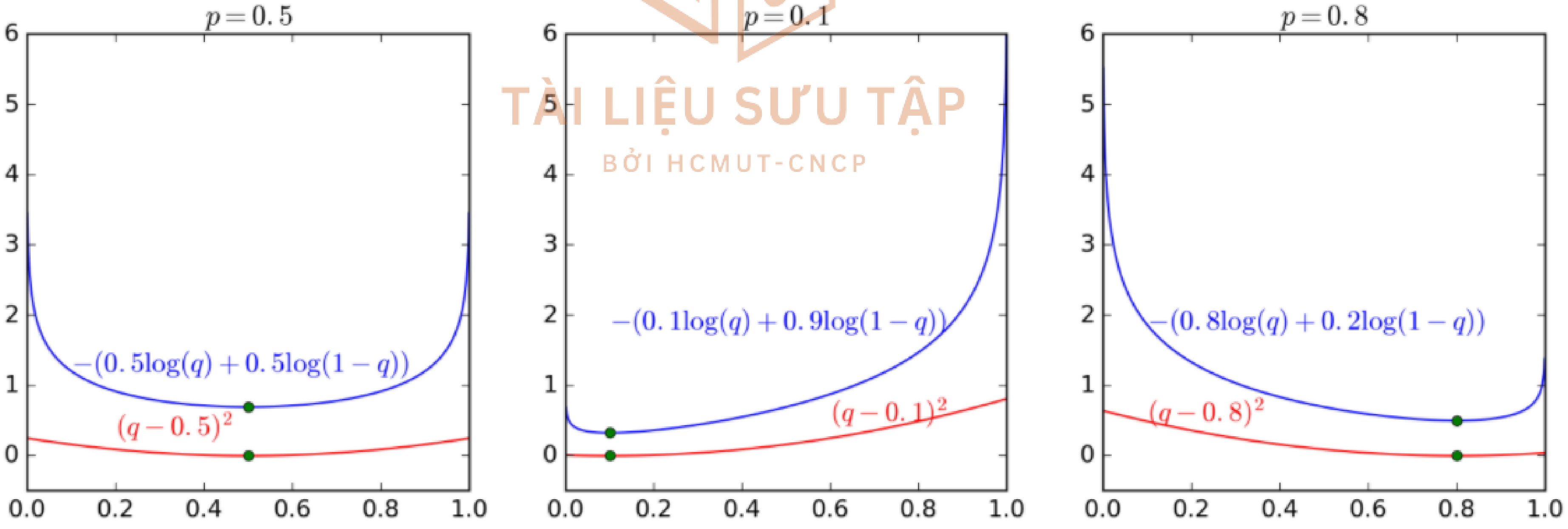
$$H(p, q) = -\sum_i p_i \log q_i$$

✓ Error back propagation

❖ Loss function:

➢ Euclidean norm (2-norm) Cross entropy

$$\|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2} \qquad H(p,q) = -\sum p_i \log q_i$$



Artificial Neural Networks

✓ Error back propagation

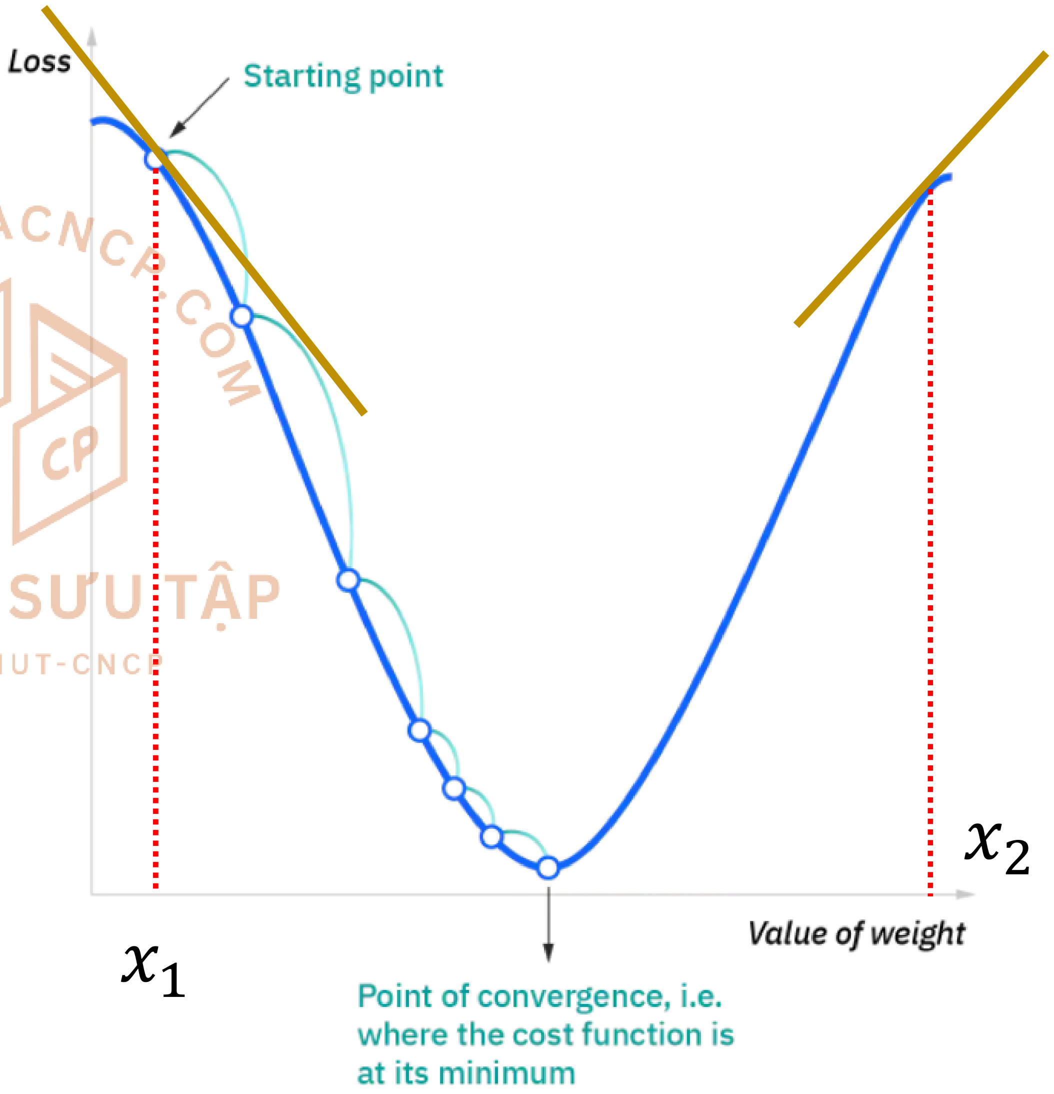
- Optimization problem
- Gradient descent

$$x \leftarrow x - \frac{df}{dx}$$

$$x \leftarrow x - \mu \frac{df}{dx}$$

μ: learning rate

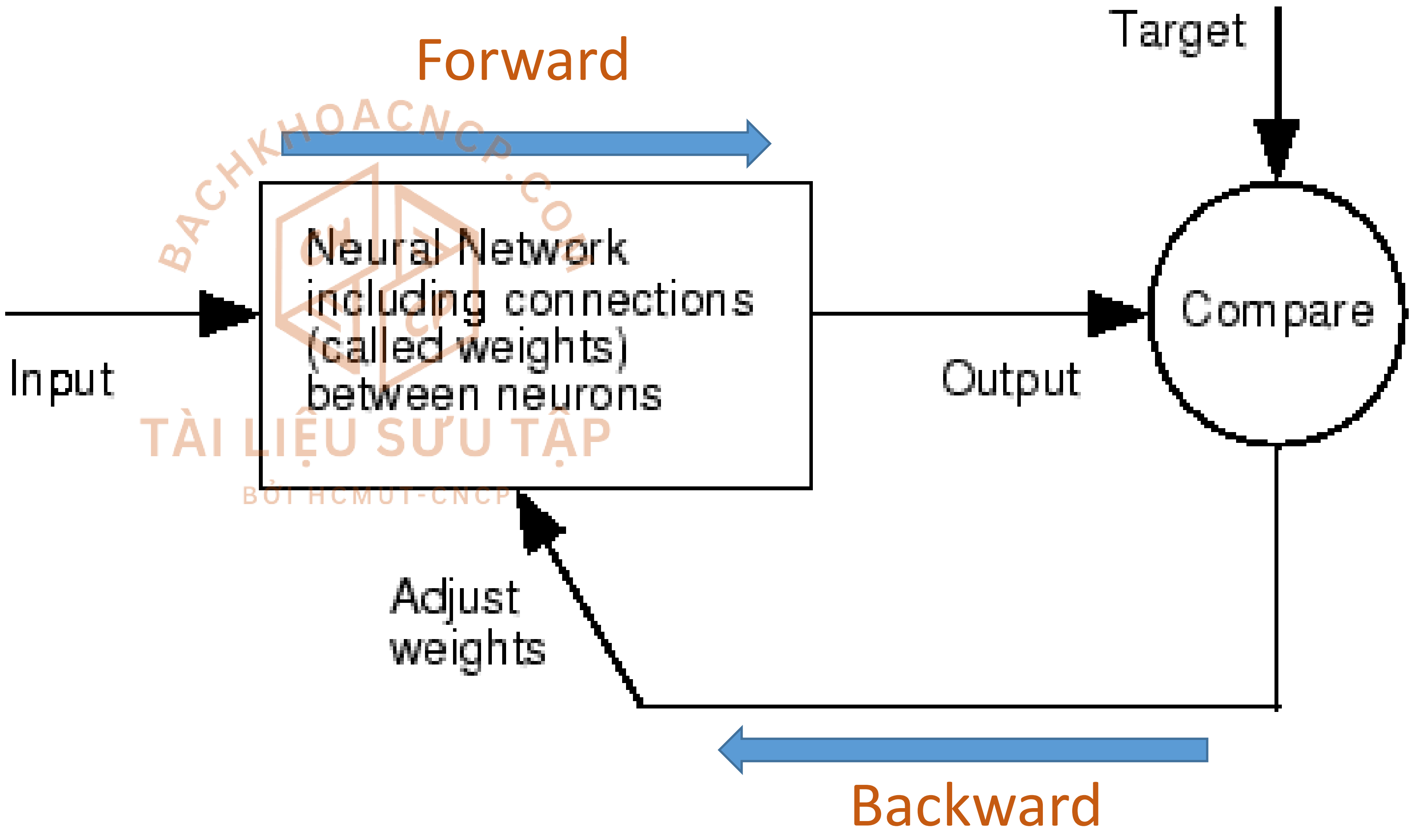
$$w \leftarrow w - \mu \frac{\partial L}{\partial w}$$



Artificial Neural Networks

✓ Error back propagation

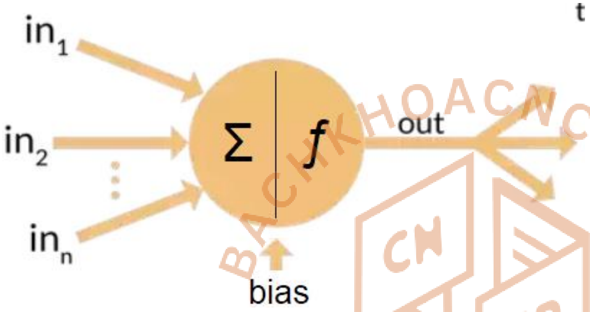
❖ Example

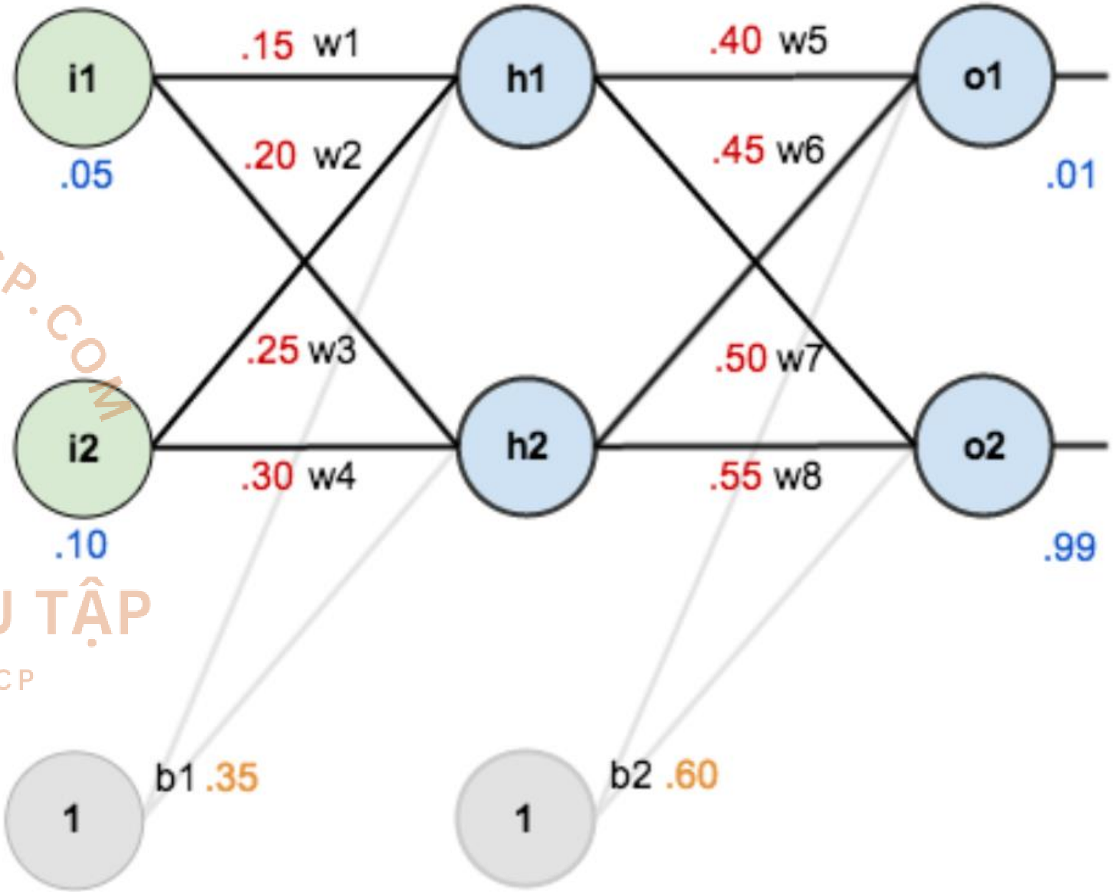


$$w \leftarrow w - \mu \frac{\partial L}{\partial w}$$

Artificial Neural Networks

- ✓ Error back propagation
  - ❖ Example





$$net_{h_1} = w_1 i_1 + w_2 i_2 + b_1$$

$$out_{h_1} = f\left(net_{h_1}\right)$$

$$f(x) = \frac{1}{1+e^{-x}}$$

$$E = \frac{1}{2}\left\{\left(target_{o_1} - out_{o_1}\right)^2 + \left(target_{o_2} - out_{o_2}\right)^2\right\}$$

Artificial Neural Networks

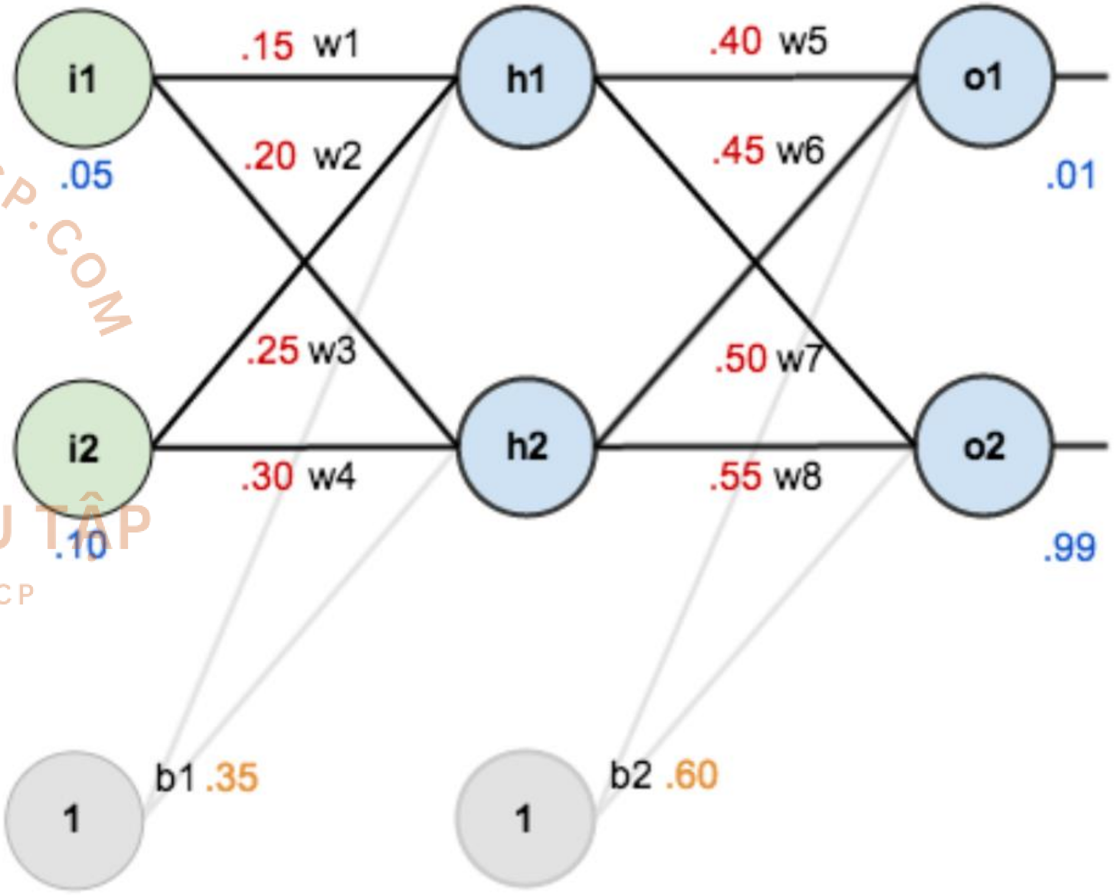
✓ Error back propagation

$$w \leftarrow w - \mu \frac{\partial L}{\partial w}$$

$$\frac{\partial E}{\partial w_5}$$

❖ Chain rule

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o_1}} \frac{\partial out_{o_1}}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial w_5}$$

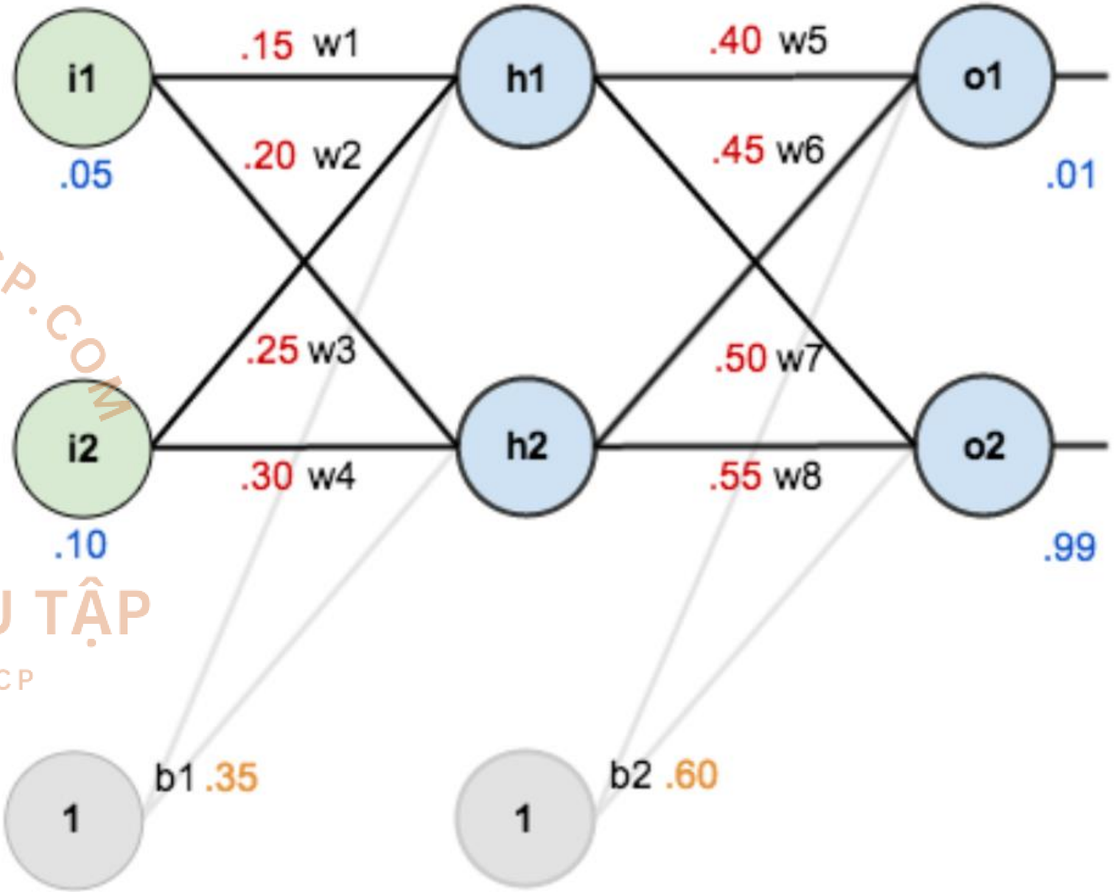


Artificial Neural Networks

✓ Error back propagation

- [Example](#)

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o_1}} \frac{\partial out_{o_1}}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial w_5}$$



$$net_{o_1} = w_5 out_{h_1} + w_6 out_{h_2} + b_2$$

$$out_{o_1} = f(net_{o_1})$$

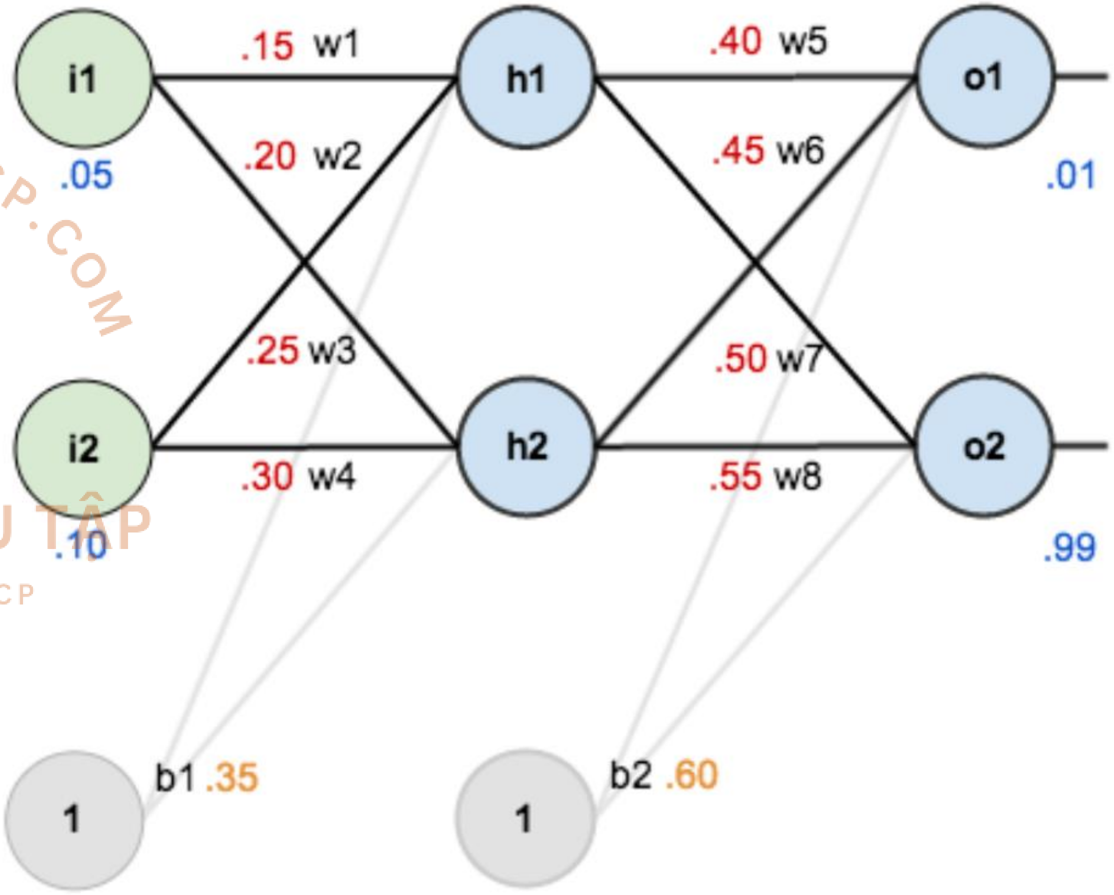
$$f(x) = \frac{1}{1+e^{-x}}$$

$$E = \frac{1}{2}\left\{\left(target_{o_1} - out_{o_1}\right)^2 + \left(target_{o_2} - out_{o_2}\right)^2\right\}$$

- Error back propagation
  - Chain rule

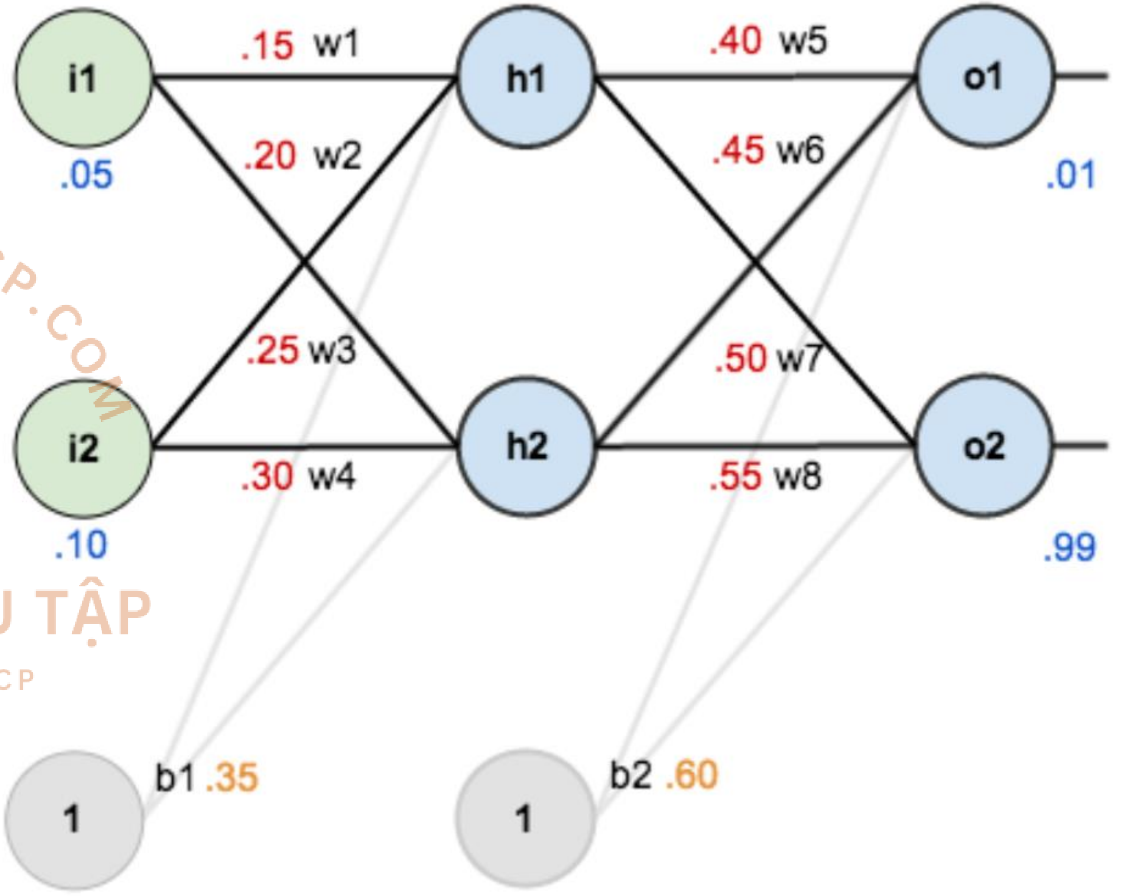
$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o_1}} \frac{\partial out_{o_1}}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial w_5}$$

w_5 → net_{O1} → out_{O1} → E



Artificial Neural Networks

- ✓ Error back propagation
  - ❖ Chain rule



$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o_1}} \frac{\partial out_{o_1}}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial w_5}$$

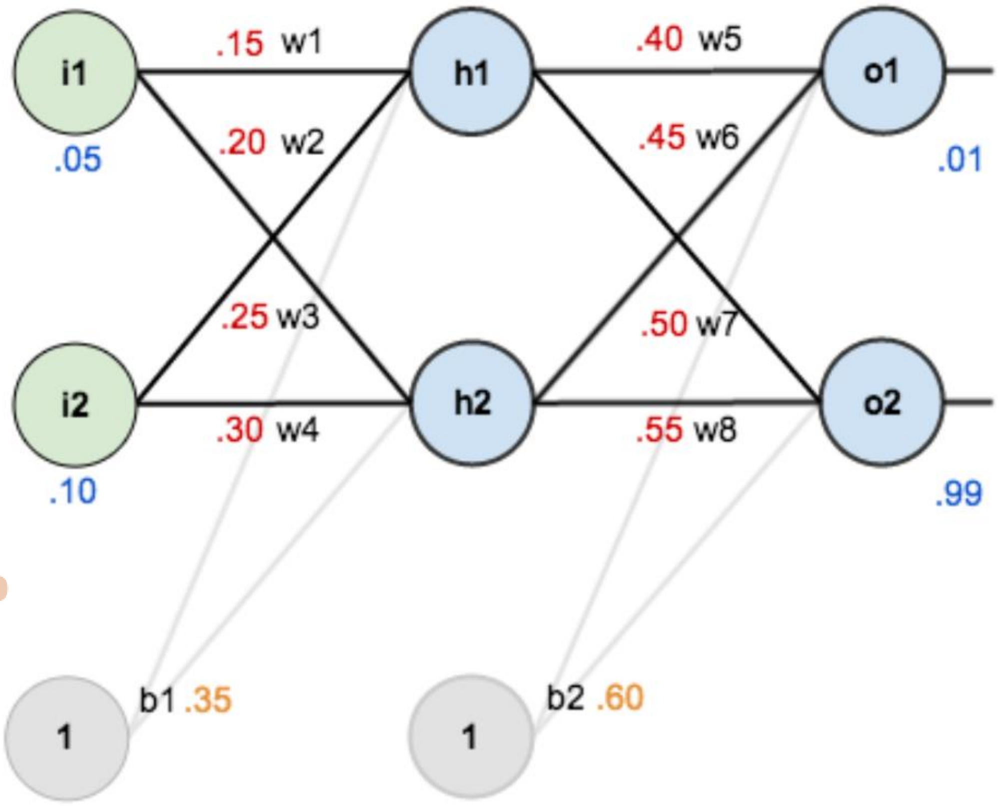
$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial out_{o_1}} \frac{\partial out_{o_1}}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial out_{h_1}} \frac{\partial out_{h_1}}{\partial net_{h_1}} \frac{\partial net_{h_1}}{\partial w_1}$$

Artificial Neural Networks

✓ Error back propagation

❖ Chain rule

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial out_{o_1}} \frac{\partial out_{o_1}}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial out_{h_1}} \frac{\partial out_{h_1}}{\partial net_{h_1}} \frac{\partial net_{h_1}}{\partial w_1}$$

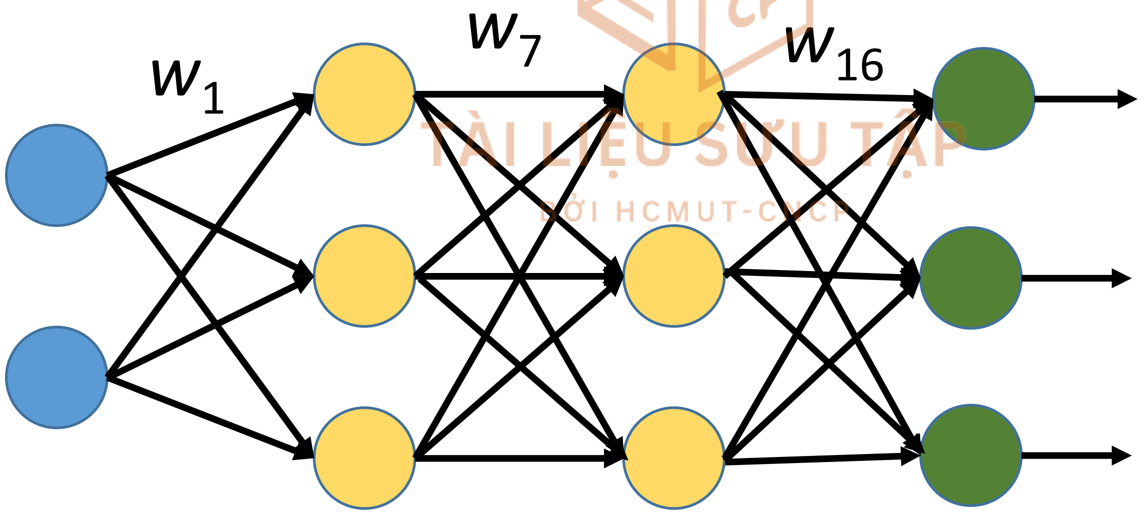


$w_1 \rightarrow net_{h1} \rightarrow out_{h1} \rightarrow net_{O1} \rightarrow out_{O1} \rightarrow E$

$out_{h1} \rightarrow net_{O2} \rightarrow out_{O2} \rightarrow E$

$$\frac{\partial E}{\partial w_1} = \left(\frac{\partial E}{\partial out_{o_1}} \frac{\partial out_{o_1}}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial out_{h_1}} + \frac{\partial E}{\partial out_{o_2}} \frac{\partial out_{o_2}}{\partial net_{o_2}} \frac{\partial net_{o_2}}{\partial out_{h_1}} \right) \frac{\partial out_{h_1}}{\partial net_{h_1}} \frac{\partial net_{h_1}}{\partial w_1}$$

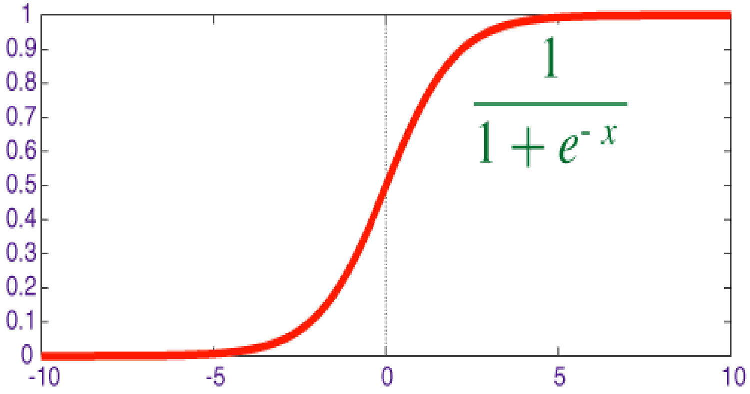
- Homework 2: Find derivative formulas for a NN
  - 2 inputs
  - 2 hidden layers, each layer has 3 neurals
  - 3 outputs



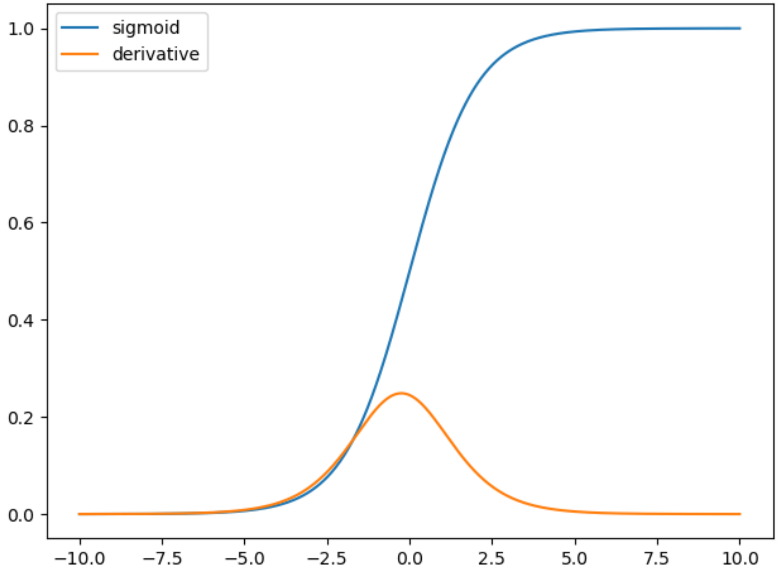
Artificial Neural Networks

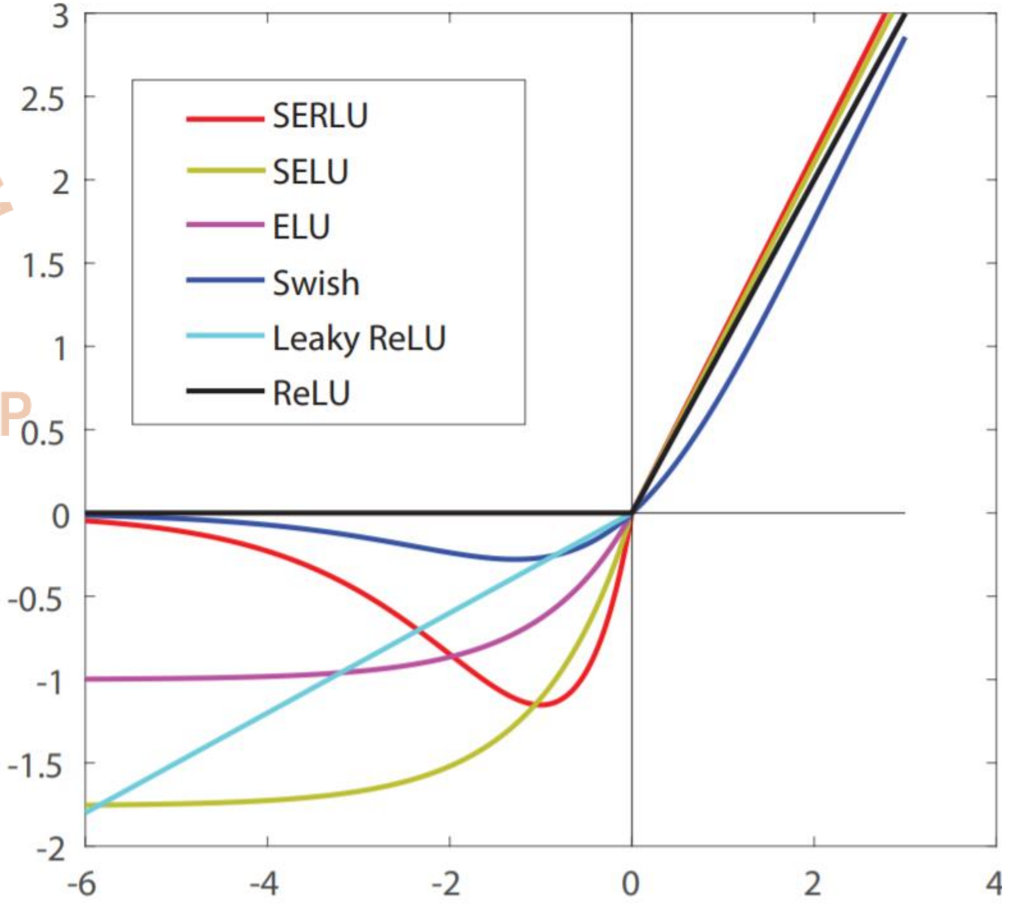
- Vanishing gradient problem

$$\frac{\partial E}{\partial w_1} = \left(\frac{\partial E}{\partial out_{o_1}}\frac{\partial out_{o_1}}{\partial net_{o_1}}\frac{\partial net_{o_1}}{\partial out_{h_1}} + \frac{\partial E}{\partial out_{o_2}}\frac{\partial out_{o_2}}{\partial net_{o_2}}\frac{\partial net_{o_2}}{\partial out_{h_1}}\right)\frac{\partial out_{h_1}}{\partial net_{h_1}}\frac{\partial net_{h_1}}{\partial w_1}$$



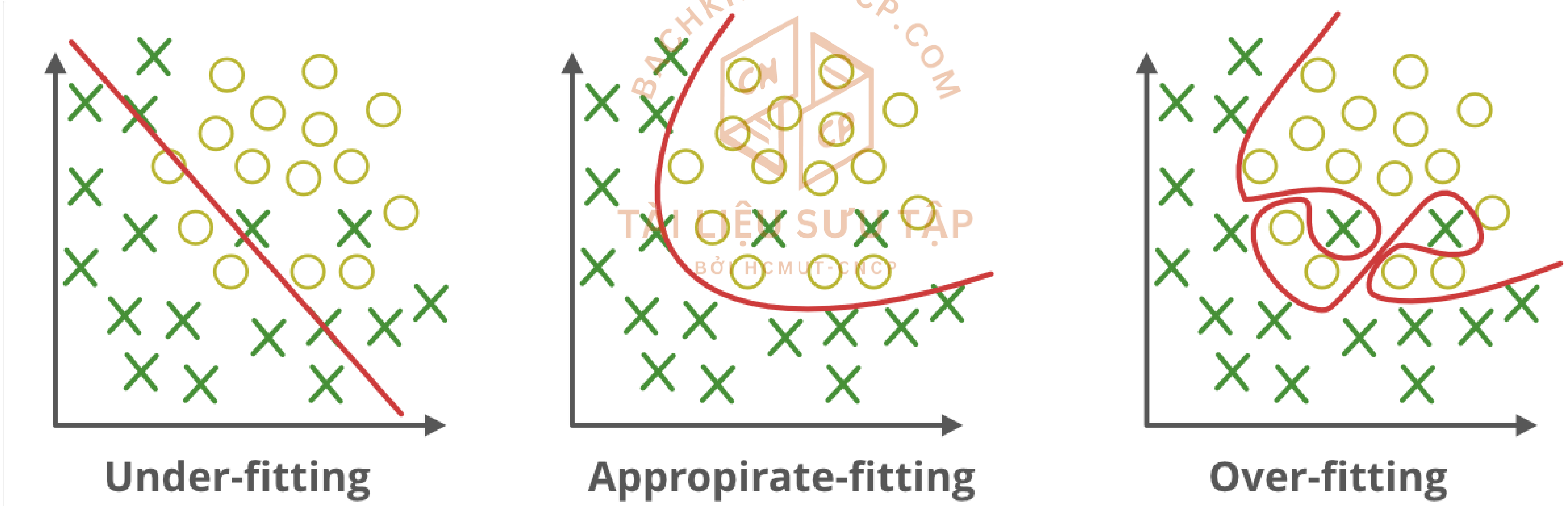
$$w \leftarrow w - \mu\frac{\partial L}{\partial w}$$





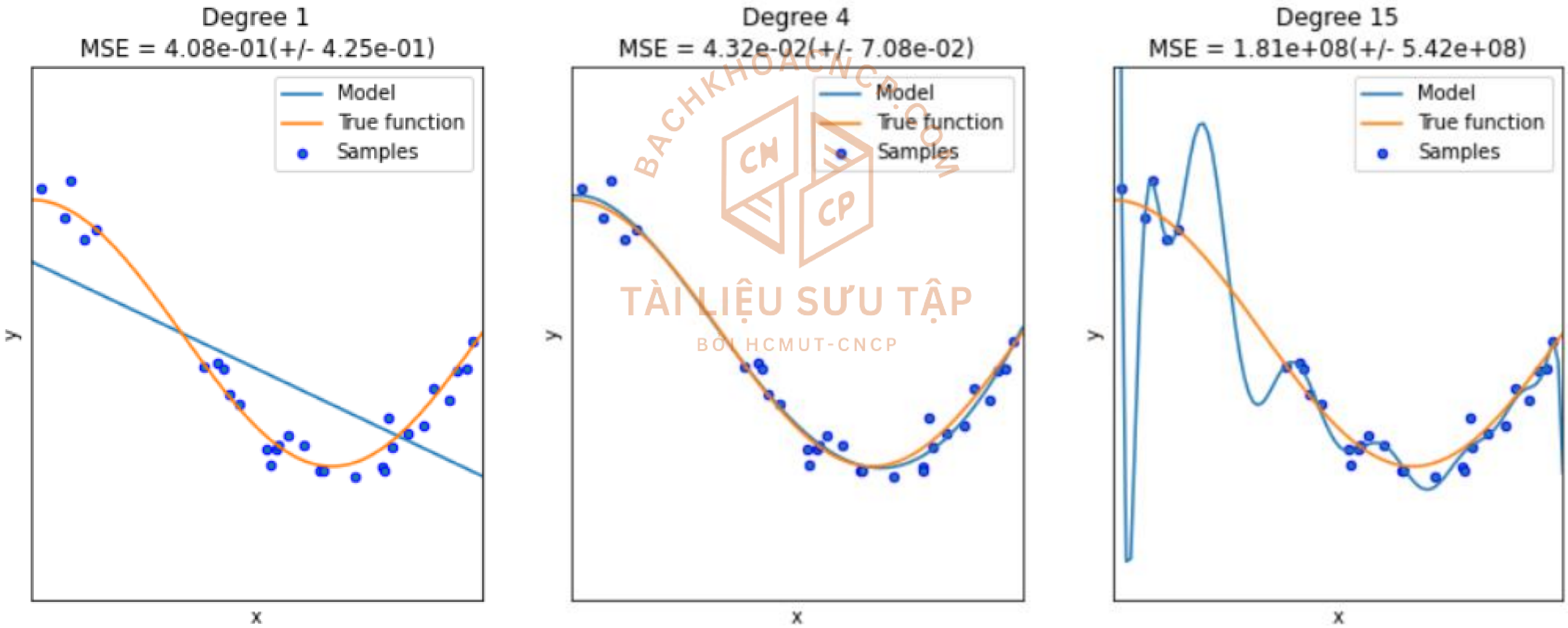
Artificial Neural Networks

- ✓ Overfitting: high variance
- ✓ Underfitting: high bias



Artificial Neural Networks

✓ Overfitting vs. Underfitting



Artificial Neural Networks

✓ Overfitting vs. Underfitting

1962	Brasil
1970	Brasil
1974	Germany
1978	Argentina
1986	Argentina
1990	Germany
1994	Brasil
2002	Brasil

✓ Overfitting vs. Underfitting

MPRA

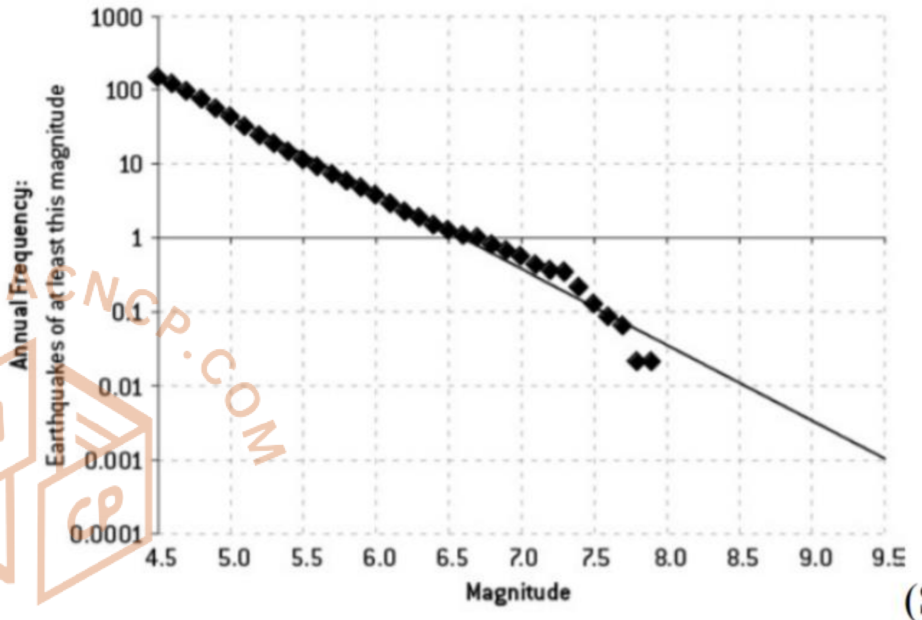
Munich Personal RePEc Archive

Fukushima: The Failure of Predictive Models

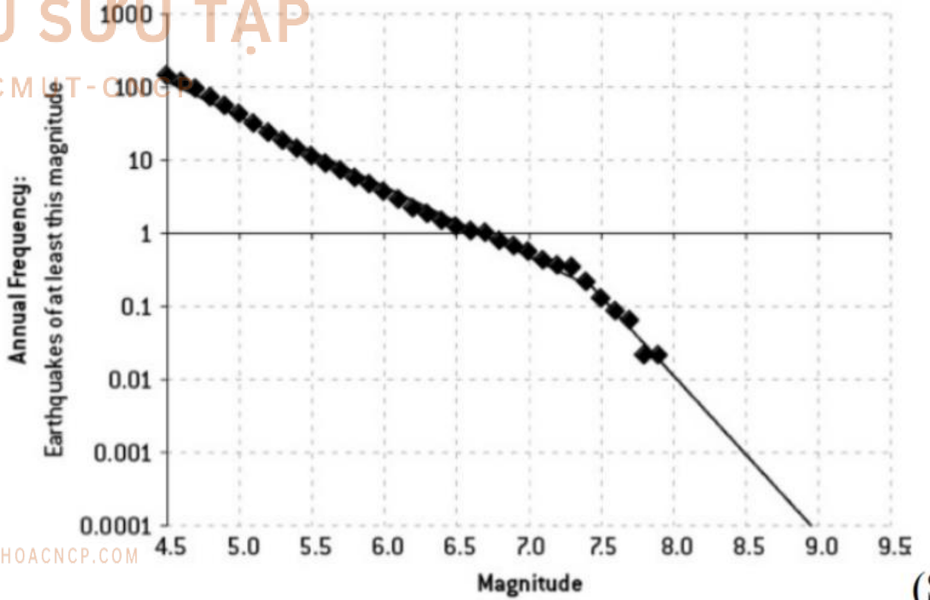
Stacey, Brian

Southern New Hampshire University

1 April 2015



(Silver, N, 2012)



(Silver, N, 2012)

Artificial Neural Networks

- Reasons
  - Underfitting (simplifying assumption)
    - Too simple model
    - Not enough training
  - Overfitting
    - Too complicated model
    - Not enough training data

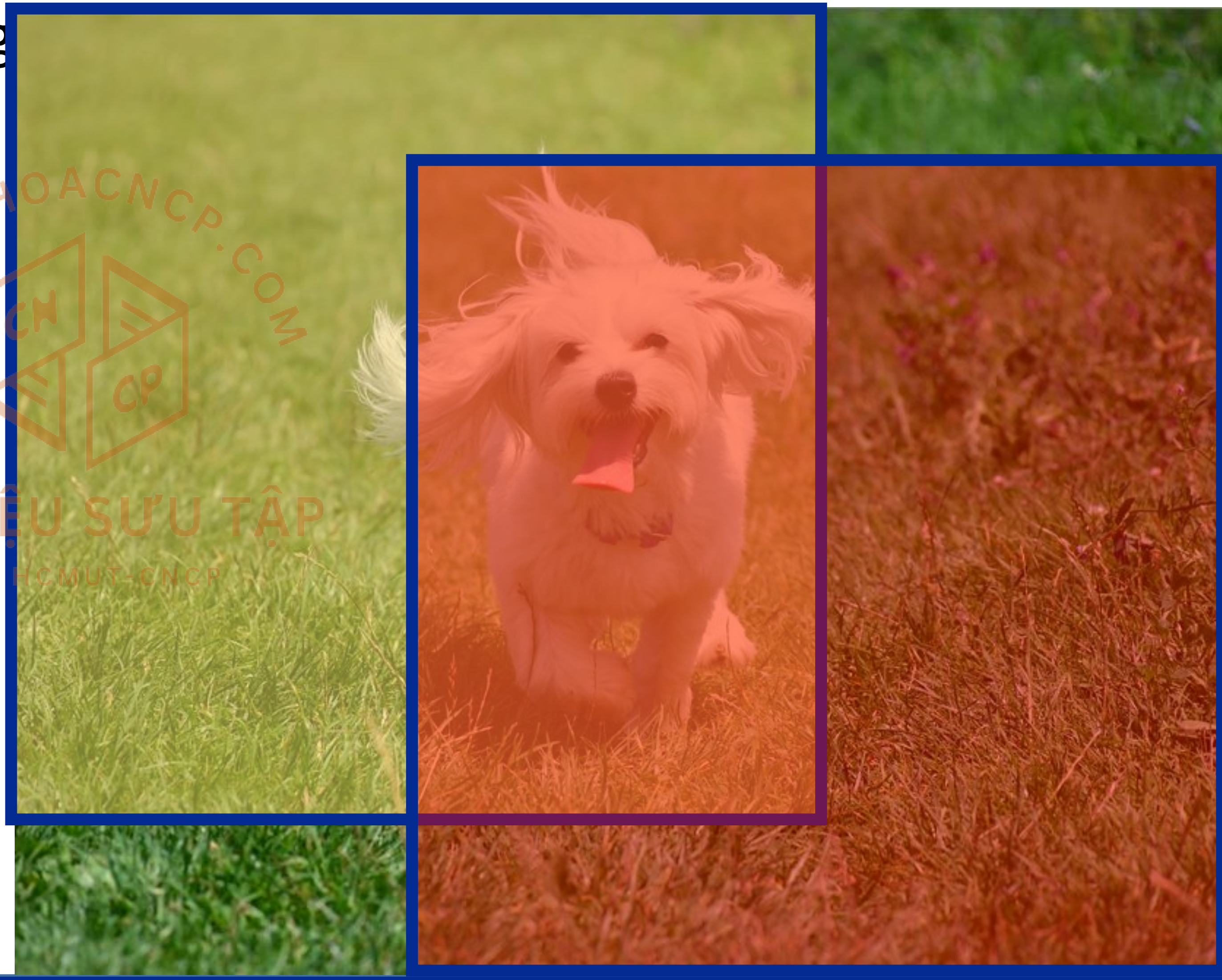
- Countermeasures for Underfitting
  - Better model
  - More training

Artificial Neural Networks

- Countermeasures for Overfitting
  - Simplifying model
  - Removing features from data
  - More data
    - ImageNet Large Scale Visual Recognition Challenge
    - 1000 classes
    - 1.2 M images

Artificial Neural Networks

- Countermeasures for Overfitting
  - Data augmentation
    - Extracting patches



Artificial Neural Networks

- Countermeasures for Overfitting
  - Data augmentation
    - Extracting patches
    - Reflection

Artificial Neural Networks

✓ Countermeasures for Overfitting

- Data augmentation
  - Extracting patches
  - Reflection
  - Rotation

Artificial Neural Networks

- Countermeasures for Overfitting
  - Data augmentation
    - Extracting patches
    - Reflection
    - Rotation
    - Altering intensities of RGB channels

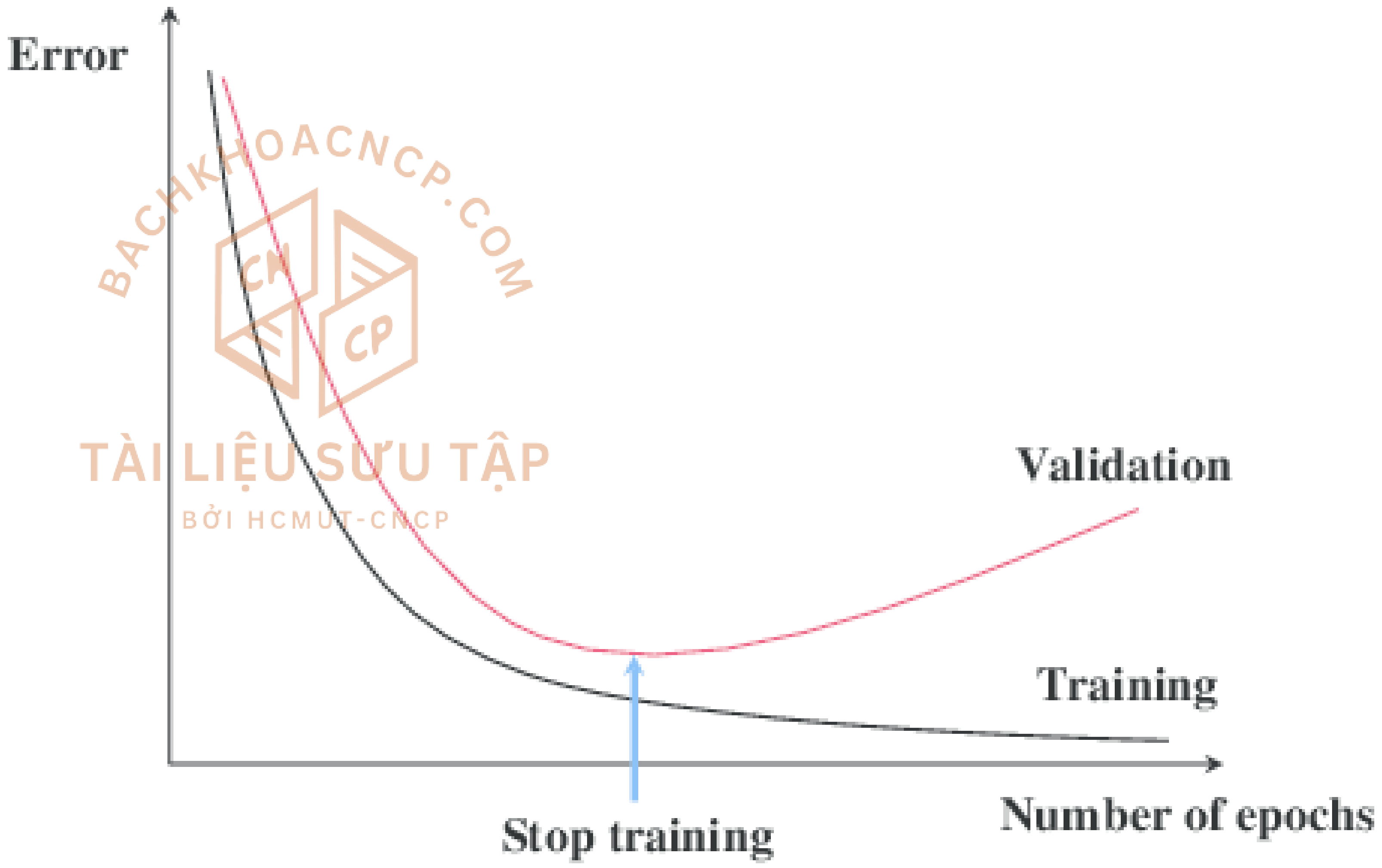
?

Artificial Neural Networks

- Countermeasures for Overfitting
  - Data augmentation
    - Extracting patches
    - Reflection
    - Rotation
    - Altering intensities of RGB channels
    - and more
    - AlexNet: 2048 times

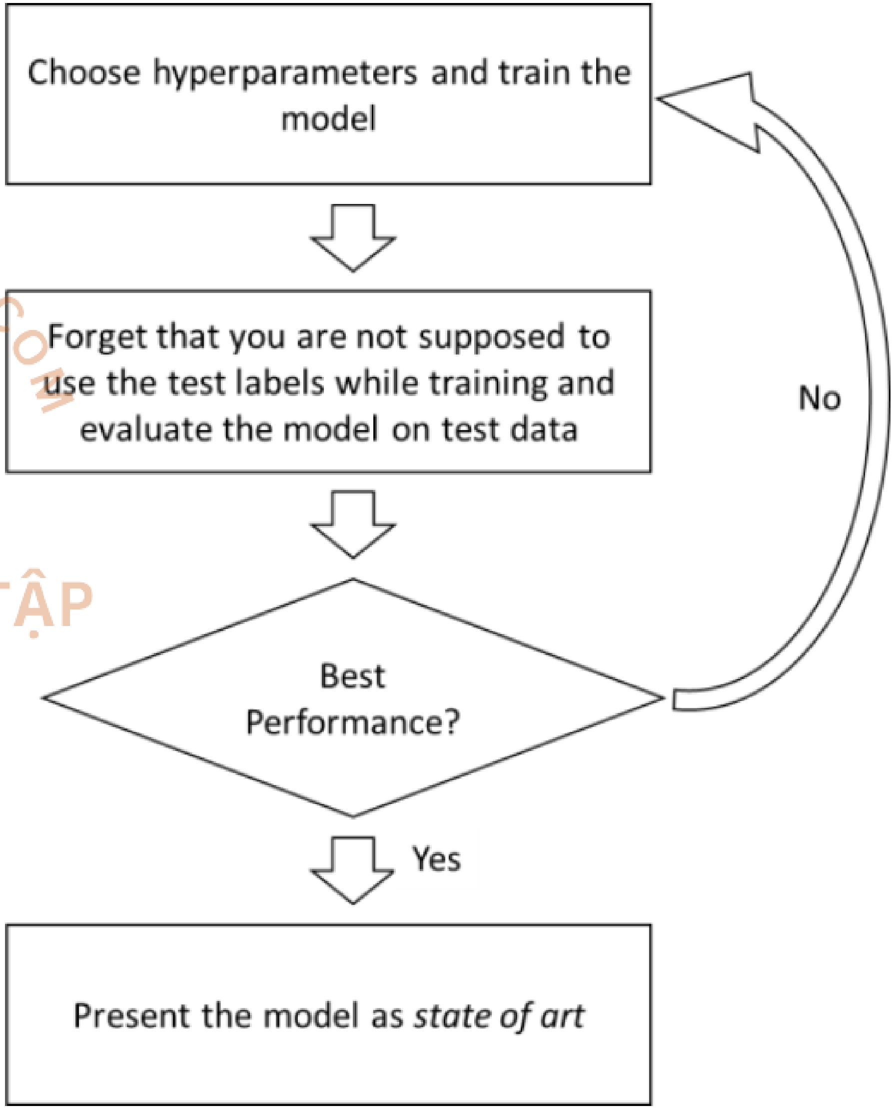
Artificial Neural Networks

- Countermeasures
  - Early stopping
    - Training set
    - Test set
    - Validation set



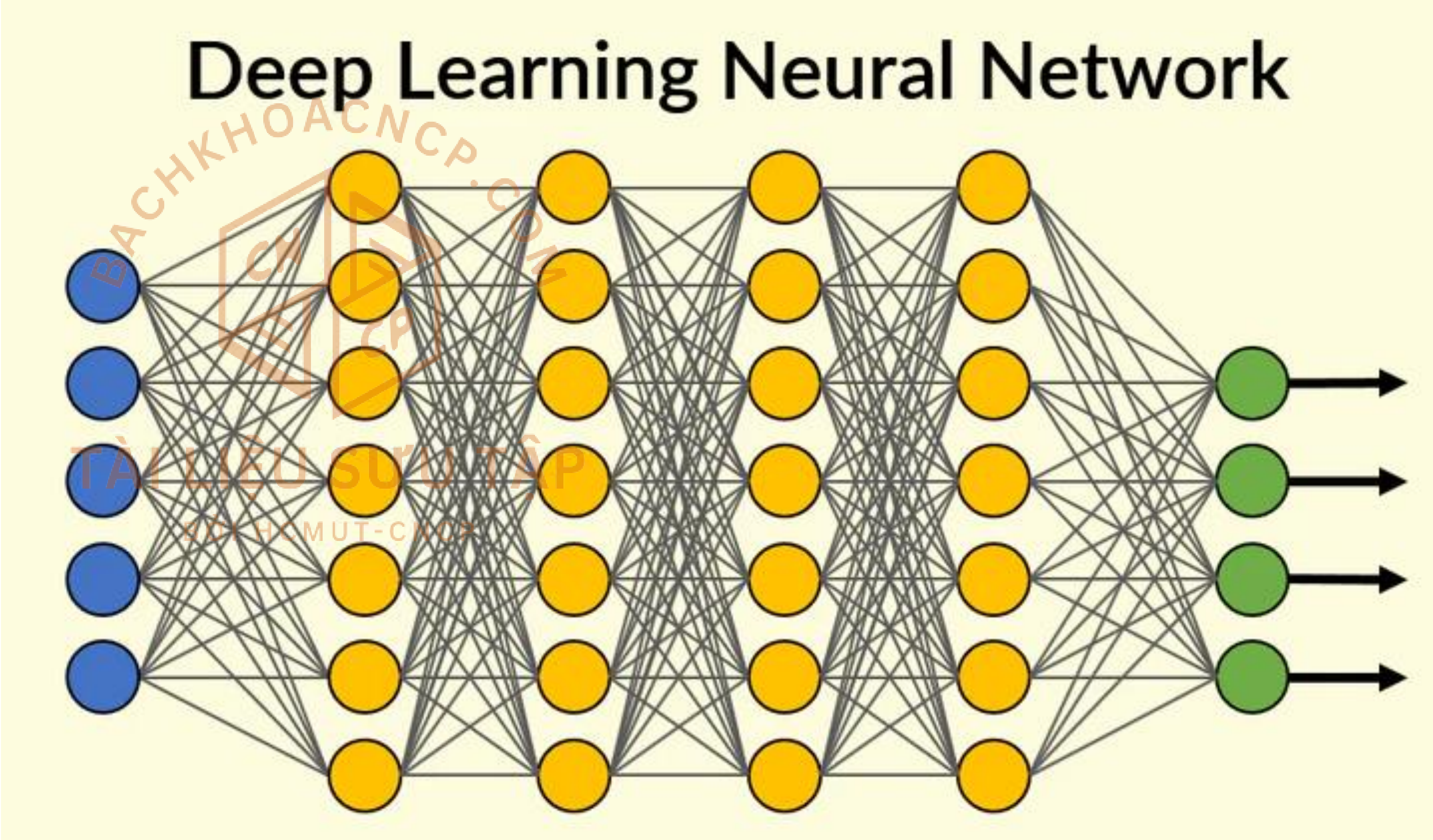
Artificial Neural Networks

- Reasons
- Countermeasures
  - Early stopping
    - Training set
    - Test set
    - Validation set



Artificial Neural Networks

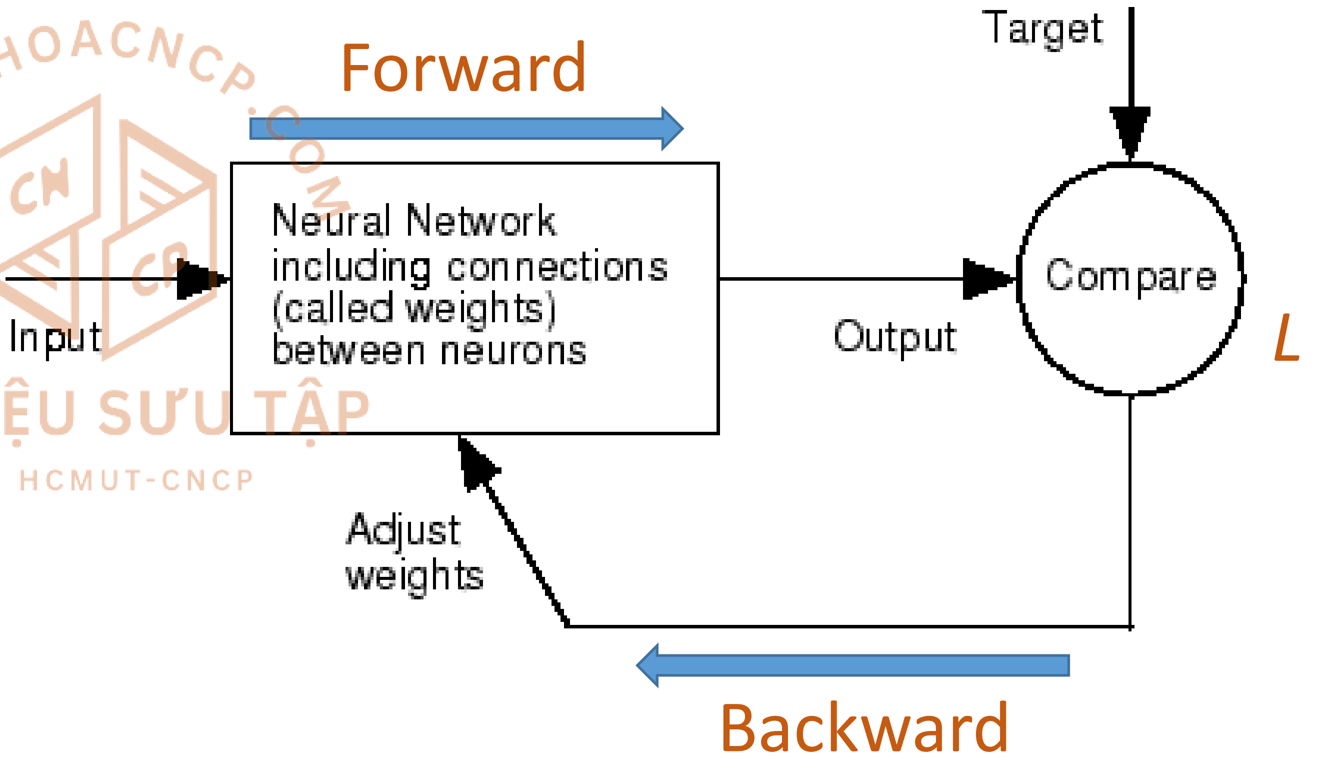
- Countermeasures
  - Drop out
    - Co-adaptation



Artificial Neural Networks

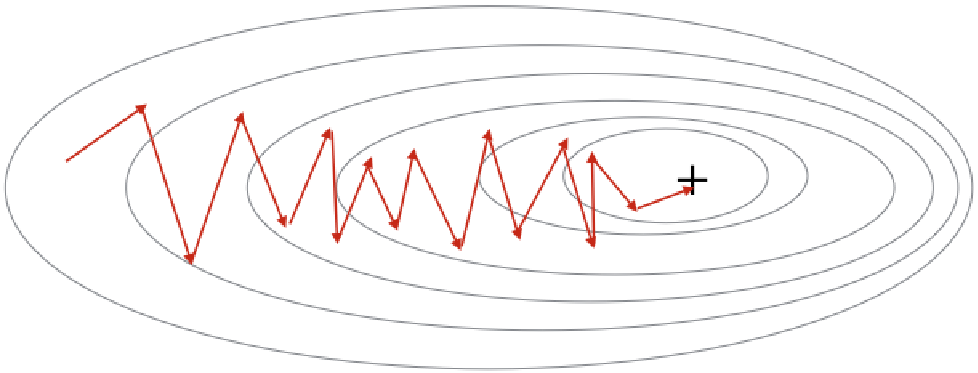
✓ Gradient descent
- 2D illustration
- Stochastic gradient descent
- Batch gradient descent

$$w \leftarrow w - \mu \frac{\partial L}{\partial w}$$

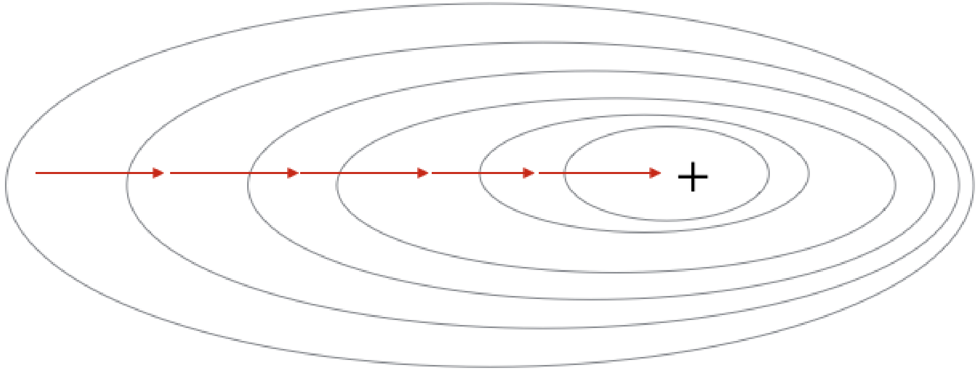


- Gradient descent
  - Stochastic gradient descent
    - Update: each training example
    - Online
    - Unstable
    - Can avoid local minima
  - Batch gradient descent
    - Update: all training dataset
    - More stable
    - May prematurely converge

Stochastic Gradient Descent

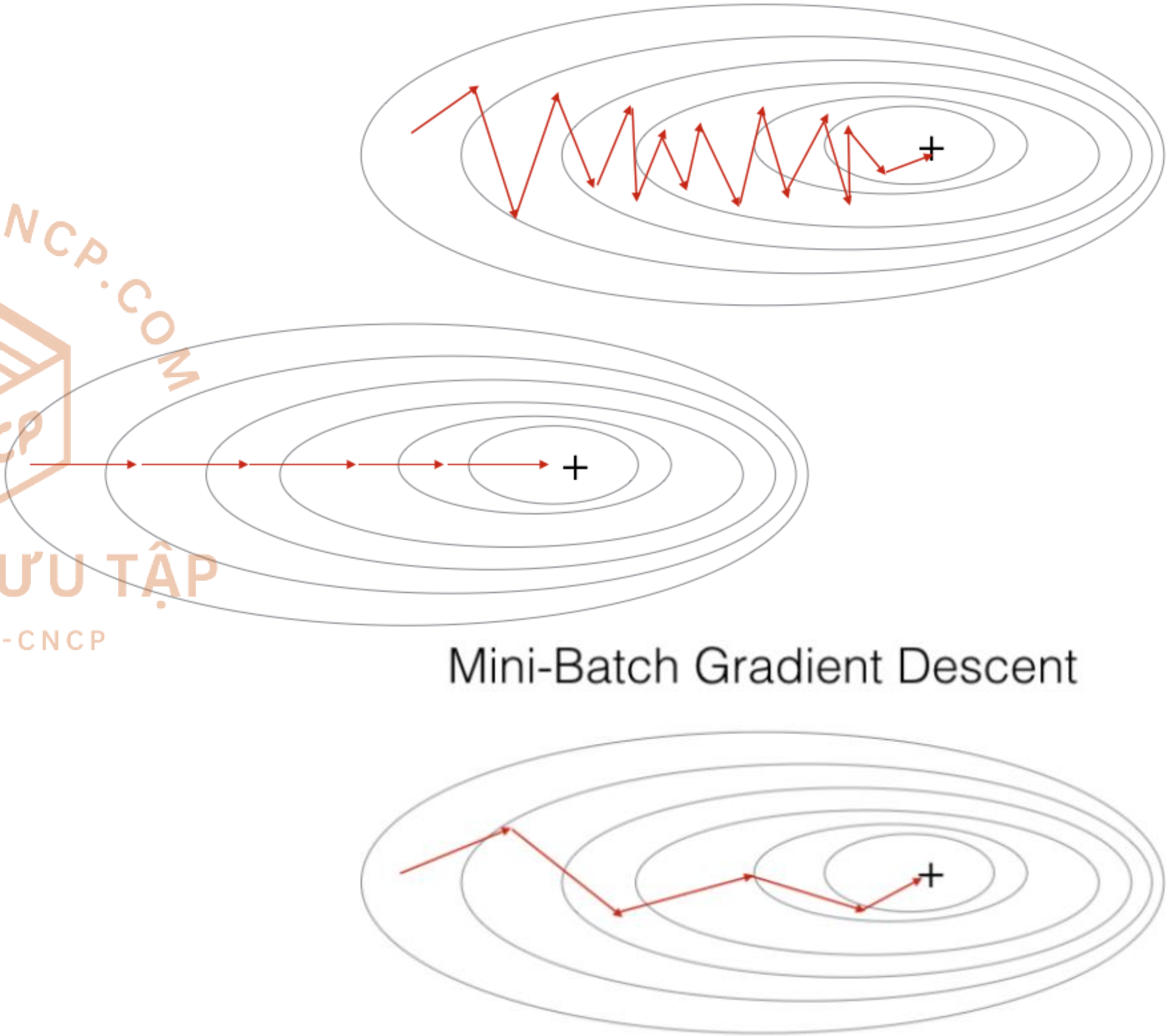


Gradient Descent



Artificial Neural Networks

- Gradient descent
  - 2D illustration
  - Stochastic gradient descent
  - Batch gradient descent
  - Mini-batch gradient descent
    - Update: subset of training set
    - Mini-batch size: 16 - 256
    - Best of both worlds



Mini-Batch Gradient Descent

Artificial Neural Networks

- Error back propagation
  - Data point/example/sample/instance/input vector/feature vector/observation
  - Batch size
  - Epoch: 10, 100, 500, 1000, and larger

Artificial Neural Networks

✓ References

- https://en.wikipedia.org/wiki/Neural_circuit
- https://en.wikipedia.org/wiki/Artificial_neural_network
- https://www.securityinfowatch.com/video-surveillance/video-analytics/article/21069937/deep-learning-to-the-rescue
- and others